New Physics: Sae Mulli,
Vol. 63, No. 9, September 2013, pp. 995∼999
DOI: 10.3938/NPSM.63.995

Stochastic Local Search in Random Constraint Satisfaction Problems

Meesoon Ha*
Graduate School of Education and Department of Physics Education, Chosun University, Gwangju 501-759, Korea

(Received 9 July 2013 : accepted 11 September 2013)

We present a method for studying the threshold behavior in random constraint satisfaction problems (CSPs) by using a stochastic local search (SLS), namely, a heuristic search. In particular, we employ the finite-size scaling concept of nonequilibrium absorbing phase transitions and address both the nature and the threshold of the solvable-unsolvable transition in terms of random K-satisfiability (K-SAT) problems, where K is the number of Boolean variables per logic clause. Based on the role of the noise parameter in the SLS, we find that the number of unsatisfied clauses (E) and the solving time ($t_{\rm sol}$) can reveal some valuable information about either the hidden structure of the solution space or the algorithmic complexity. As compared to two-value averaging over different samples, we show that survival-sample-averaged quantities in the steady-state limit are good and clear indicators of both the nature and the threshold of the phase transition in the thermodynamic limit.

PACS numbers: 05.40.-a, 02.70.-c, 64.60.Ht, 89.20.Ef
Keywords: Heuristic, Random constraint, K-SAT, Threshold behavior, Noise, Finite-size scaling

I. INTRODUCTION

Among constraint satisfiability problems (CSPs), the K-satisfiability (K-SAT) problem has been well-known in computer science as one of the famous non-deterministic polynomial (NP)-complete class problem. Recently, statistical physics has entered this field with its ability to handle the complicated energy landscape and frustration from the quenched disorder, which are analogous to the glassy or spin-glass systems [1–6]. The central question of the K-SAT problem is finding out of a configuration of N Boolean variables *satisfying* an *instance* F, which is a Boolean expression: $F = \Big[C_1 \text{ AND } C_2 \text{ AND } \cdots \text{ AND } C_M\Big]$, where each *clause* C_i is given by $C_i = (y_{i1} \text{ OR } y_{i2} \cdots \text{ OR } y_{iK})$ and each y_{ij} is randomly picked from the set $\{x_1, x_2, ..., x_N\}$ of N Boolean variables themselves or their negation.

The Boolean expression characterized by the number of clauses M and the number of variables per clause K in this fashion is called the conjunctive normal form (CNF), where the density of constraint $\alpha \equiv M/N$ play a role as a control parameter that determines the problem to be satisfiable (SAT) or unsatisfiable (UNSAT) [2]. In numerics, the maximum running time $T_{\rm max}$ is another control parameter. As α increases, it gets harder to find the SAT configuration of variables and finally the solution does not exist for too large values of α. The *critical* value of α, denoted as α_c, is very similar to the critical point of the phase transitions in statistical mechanics.

Since the number of possible configurations for N Boolean variables is 2^N, it is not plausible to search the entire configuration space with the brute force. Solving techniques developed so far classified as two kinds, which are deterministic [2,8,9] and stochastic [10–13] ones.

Stochastic-local-search (SLS) algorithms basically flip a trial variable and check the status of the expression after flippling, *e.g.*, the number of UNSAT clauses, to determine whether the flipping is accepted or not. If we consider the number of UNSAT clauses as some kind of *energy*, the local search heuristics are reminded of Monte Carlo simulations [14,15], familiar to physicists studying phase transitions. In fact, the focused Metropolis search

*E-mail: msha@chosun.ac.kr

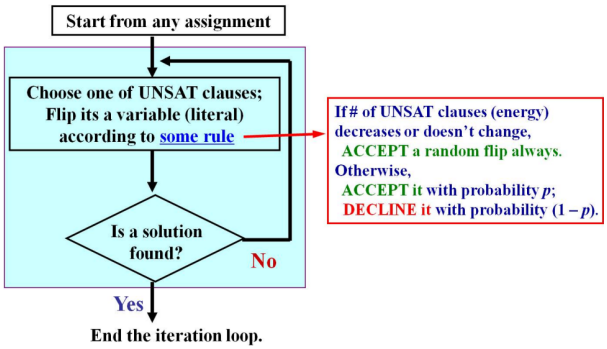

Fig. 1. (Color online) Flowchart of the SLS algorithm, ASAT for random K-SAT problems.

(FMS) algorithm [11] employs the exactly same mechanism with the standard Metropolis algorithm with the number of unsatisfied clauses as energy.

In this paper, we briefly discuss how to analyze SLS heuristics, which is simple and fruitful for beginners as discussed in our earlier work [16], where we presented how to characterize random K-SAT problems, in terms of the standard theory of phase transitions, especially finite-size effects.

As an example, we take the average SAT (ASAT) [12] incorporating a noise parameter p, which is a simplified version of FMS, and show various aspects of random CSPs, in the context of the solvability (satisfiability) of random K-SAT with energy (the number of remaining UNSAT clauses) and the solving time after the test time limit of the algorithm denoted as $T_{\max}$. Our numerical analysis is applied to distinguish the different transition nature, which depends on the value of K [6].

This paper is organized after introduction as follows: In Sec. II, we briefly review the ASAT algorithm and measure relevant physical quantities near the threshold against two different sampling methods. Based on the finite-size scaling concept of nonequilibrium absorbing phase transitions, we give some naive picture for the solvability of random K-SAT problems against α and p, as well as the optimal value of p where the threshold α_c of the solvability and the transition nature are meaningful. Finally, possible applications in random CSPs are discussed in Sec. III with graphical snapshots in the network (factor-graph) representation and open questions.

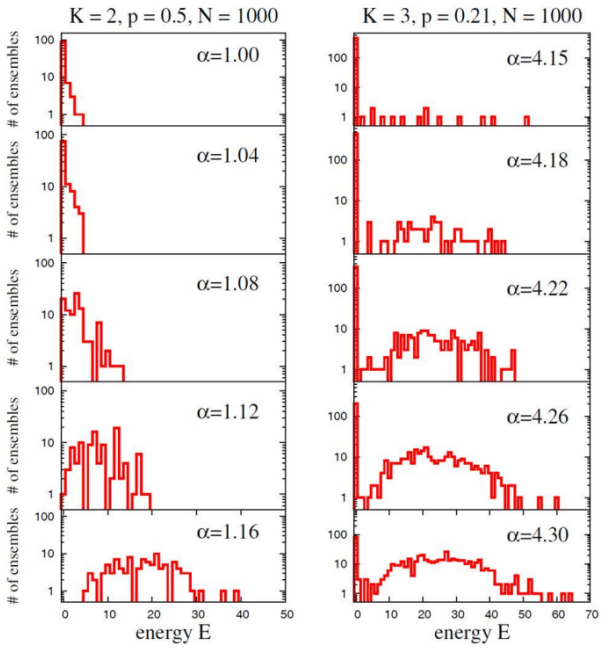

Fig. 2. (Color online) As α varies, energy histograms are shown for $K = 2$ ($N = 16000$ at $p = 0.5$) with 100 samples in the left panel; for $K = 3$ ($N = 1000$ at $p = 0.21$) with 500 samples in the right panel.

II. NUMERICAL RESULTS

Figure 1 is the pseudo code of the SLS algorithm, for random K-SAT problems, where some rule can be a trial movement of a variable in an UNSAT clause is accepted if it reduces the energy E (the number of UNSAT clauses) or with probability p (the *noise* parameter), if it increases the energy. It is known that the *optimal value* of p ($p_{\rm opt}$), the value poised between too less noise to prevent the system from escaping from the local energy minima and too much fluctuations, is 0.21 for $K = 3$ reported in [12], with which the ASAT algorithm finds a solution the fastest and up to the largest value of α. In order to make this clear, we measure time to solve random K-SAT problems in the unit of flips/N, as in the previous works. The energy E, defined as the number of UNSAT clauses at the end of each simulation, is normalized by dividing it with the total number of clauses M and considered as "energy density."

For the different transition nature reported earlier [6], we show how to distinguish it with energy histograms. Figure 2 represents the different characteristics in energy histograms for $K = 2$ (smoothly moving distributions as α increases) and $K = 3$ (abrupt change of distributions

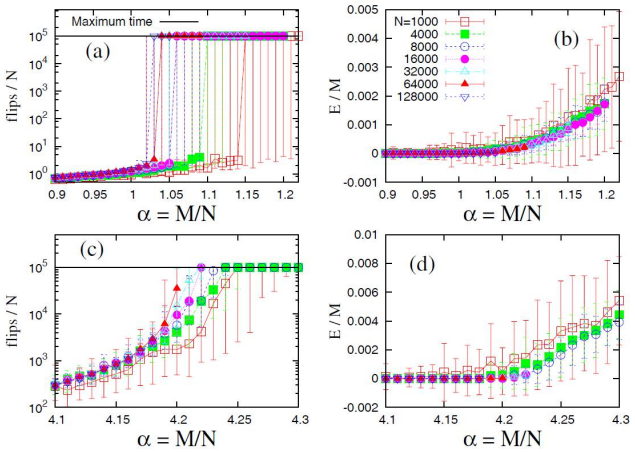

Fig. 3. (Color online) Simulation results of ASAT for random K-SAT problems at $p_{\rm opt}(K)$: $t_{\rm sol}$ (the number of flips/N) and energy density (E/M) are plotted as (a) and (b) for $K = 2$ at $p = 0.5$; (c) and (d) for $K = 3$ at $p = 0.21$. Symbols and error bars indicate the median and quartiles for (a) and (c), averages and standard deviations for (b) and (d), where 100 samples are taken.

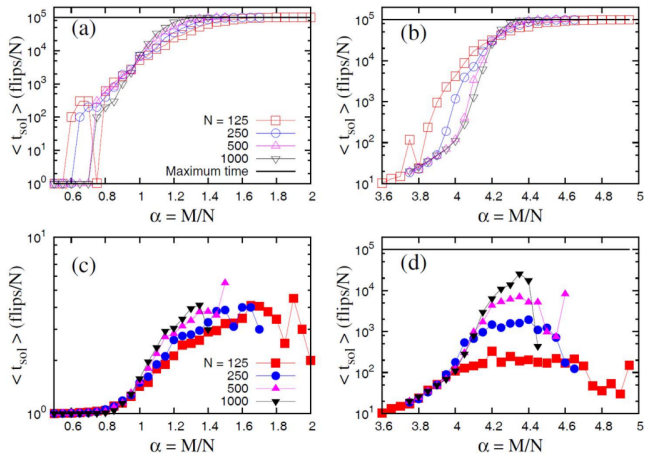

Fig. 4. (Color online) The solving time $t_{\rm sol}$ against α for various N: (a) and (b) for $K = 2$ are averaged over survival/active (unsolved, UNSOL) samples; (c) and (d) for $K = 3$ are averaged over dead (solved, SOL) samples. We set $T_{\rm max} = 10^5$ and take 1000 samples.

as α increases or phase coexistence), respectively. This implies that the transition for $K = 2$ is continuous, but discontinuous for $K = 3$, just as discussed in Ref. [6]. If one extends this issue to $(K + P)$-SAT problems, where the fraction P of the clauses has K variables and the rest of clauses has $K+1$ variables (on average, $(K+P)$ variables per clause), the distinction of transition natures is located at $K_* \simeq 2.41$ [6]. We confirm that energy histograms for the case of $K = 2.2$ show the similar behavior to the case $K = 2$ (continuous), and those for the case

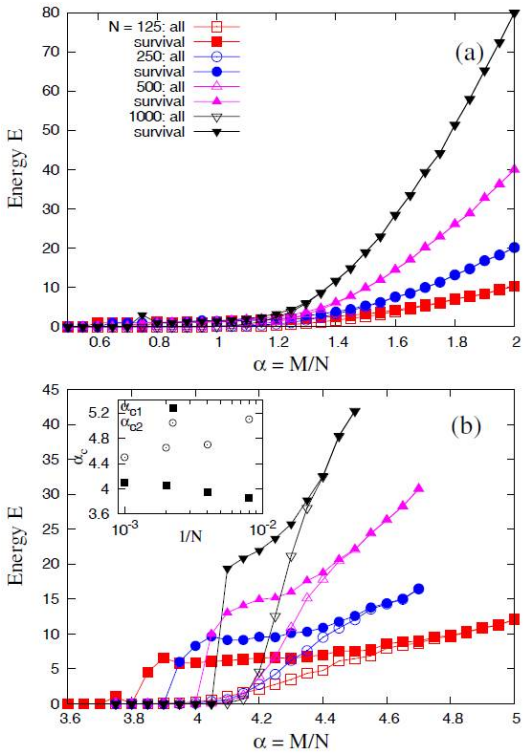

Fig. 5. (Color online) Energy E against α for various N: (a) for $K = 2$ and (b) for $K = 3$, where open (solid) symbols are averaged over all (survival/active/UNSOL) samples. All other setups are exactly the same as Fig. 4. In the inset of (b), we plot two different locations of phase boundary threshold, denoted as α_{c1} and α_{c2}.

$K = 2.6$ looks similar to the case $K = 3$ (discontinuous), as expected.

Figure 3 shows numerical results for random K-SAT problems at $p_{\rm opt}(K)$: $p_{\rm opt}(3) \simeq 0.21$ [12] and $p_{\rm opt}(2.6) \simeq 0.30$ obtained from our numerics. Due to the fact that there is no dependence of p for $K < K_*$, we take the arbitrary middle value $p_{\rm opt}(K < K_*) = 0.50$. Interestingly, the time to solve random K-SAT problems abruptly changes for $K = 2$ and smoothly changes for $K = 3$, in contrast to the intuitive and previously presented results of energy for $K < K_*$ (continuous) and for $K > K_*$ (discontinuous). Therefore, we find that the scaling behavior of energy and time should be carefully treated, in the context of the finite-size scaling (FSS) analysis.

Based on the FSS analysis, the results of Fig. 4 for $\langle t_{\rm sol} \rangle_{\rm SOL}$ can indicate threshold from the SOL phase to the UNSOL one, whereas the results of Fig. 5 for $\langle E \rangle_{\rm UNSOL}$ can do so from the UNSOL phase to the SOL one as a continuous transition. The detailed FSS analysis can be found in [16]. In the theory of critical phenomena,

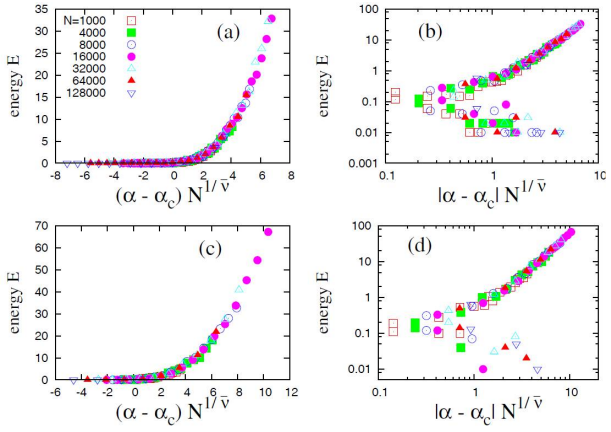

Fig. 6. (Color online) FSS of $\langle E\rangle_{\rm all}$: for $K=2$ at $p=0.5$, (a) linear scales and (b) double-logarithmic scales with ($\alpha_c = 1.00$, $\bar{\nu} = 2.75$); for $K=2.2$ at $p=0.5$, (c) as (a) and (d) as (b) with ($\alpha_c = 1.25$, $\bar{\nu} = 2.60$). Each data point in both cases is averaged over 100 samples.

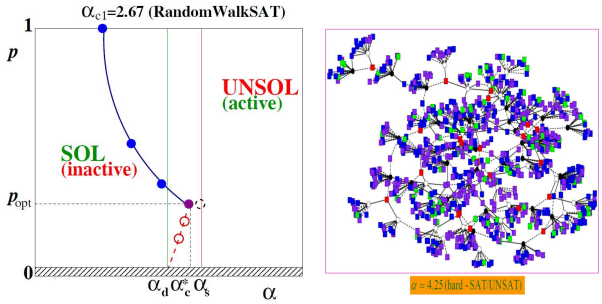

Fig. 7. (Color online) Schematic phase diagram of ASAT for $K=3$ with numerical data points: The $p=1$ is the same as RandomWalkSAT [10]. At $p=0$, because of the trapping even at minimal height of energy barrier, it is extremely difficult to determine any transitions. The threshold point α_c^* represents the SOL-UNSOL transition at $p_{\rm opt}=0.21$. Two vertical lines indicate the SAT-UNSAT transition $\alpha_s \simeq 4.267$ [17] and the clustering (dynamic) transition of solutions $\alpha_d \simeq 3.86$ [18], respectively. Note that α_c^* is located between α_d and α_s in the "hard" SAT region (left); the complexity of the threshold at $p=p_{\rm opt}$ (right).

the correlation volume diverges near the critical point as $\xi_{\rm v} \sim \epsilon^{-\bar{\nu}}$, where ϵ is the reduced control parameter and $\bar{\nu}$ is called the FSS exponent. In the context of random K-SAT problems, $\epsilon \equiv (\alpha-\alpha_c)/\alpha_c$, and the proper scaling variable x is given by $x \equiv \epsilon N^{1/\bar{\nu}} \propto (\alpha-\alpha_c)N^{1/\bar{\nu}}$.

Figure 6 shows the FSS collapse of E using the FSS exponent $\bar{\nu}$. The quality of data collapse is excellent for $K=2$ and $K=2.2$, and the values are compatible with those the reported previously with some different approach [6], while the quality of data collapse is not quite good for $K=2.6$ and 3 (data not shown) if all sample averaged data are collapsed just as the case of $K<K_*$. However, the value of $\bar{\nu}$ tends to decreases as K increases, which is also the same as the earlier results [6]. So we suspect that the discontinuous nature of transitions for $K \gtrsim K_*$ is related to the relatively poor quality of FSS collapse due to the nontrivial sampling issue and the finite-size correction to scaling.

The sampling issue in random K-SAT problems using SLS heuristics is important. The solution of the problem can be represented as the ground state energy that is identically zero, while the number of solutions, represented as entropy, increases as α decreases [5]. In terms of the absorbing phase transition (APT) in nonequilibrium processes, a ground state corresponds to an absorbing state, *i.e.*, UNSOL instances can be interpreted as the "survival (active) samples" in APTs. Therefore, it is very useful to distinctively deal with instances with nonzero energy values (at the end of the simulation/search due to the maximum time of search) and to consider averaging only over UNSOL instances as the standard notion of the survival sampling average in APTs. In case of $K \gtrsim K_*$ where the phase coexistence occurs near α_c, such distinction plays a crucial role in the FSS analysis for the threshold behavior. The characteristic jump of the energy value for $K=3$ is indeed observed as shown in Fig. 5, when the survival sample average is used, which reveals the discontinuous transition along with energy histograms in Fig. 2. Note that in Fig. 3 where the values are averaged over all samples, the jump is hidden.

Previously, E is defined as the number of *leftover* UNSAT clauses at the end of each simulation. However, the number of UNSAT clauses *during* the simulation can also reveal the structure of solution spaces or the properties of specific algorithm used. Thus, we need to monitor the change of energy, $\Delta E(t)$ during ASAT simulations, and analyze the temporal scaling behavior of E. As discussed in [16], the instances for $K<K_*$ actually are separated quickly into SOL and UNSOL ones, while the separation comes much later for the instances for $K \gtrsim K_*$. Moreover, in some cases, $E(t)$ does not "saturate" until $T_{\rm max}$. As a result, the saturated value of $E(t)$ may depend of the choice of $T_{\rm max}$, but we confirm that the different character of histograms for $K=2$ and $K=3$, shown in Fig. 2, is still valid against the value of $T_{\rm max}$.

III. SUMMARY AND REMARK

In summary, we have demonstrated how to analyze random K-SAT problems using SLS heuristics in terms of a similar numerical framework to that used to investigate nonequilibrium processes. Although these problems have been intensively studied in physics community due to the characteristic frustration and complicated energy landscape, it is a bit surprising that finite-size effects have not been systematically investigated is in the language of critical phenomena, except for a few cases where the FSS exponent $\bar{\nu}$ is considered [2,5,6]. With $p_{\rm opt}(K)$ of ASAT, we have focused on the transition occurred with the variation of $\alpha = M/N$, and confirmed the different transition nature below and above $K = K_* \simeq 2.41$ based on the FSS of energy with the proper sampling. As sketched in Fig. 7, it is still an open question for the possibility of two different transitions for $K \gtrsim K_*$ and the validity of the FSS exponent. We also note that recent series of researches [19, 20] reveal that the complicated nature of the transition for $K \geq 3$ comes from the "mixture" of instances showing continuous and discontinuous transitions. Dividing instances into such different domains can be another good strategy for future works.

ACKNOWLEDGMENTS

Many of the results and the ideas presented here grew out of an enjoyable collaboration and fruitful discussion with Sang Hoon Lee in the early stage of our random K-SAT study (see Ref. 16). This work is supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (NRF-2011-0011550).

REFERENCES

[1] M. Mézard, G. Parisi and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).

[2] S. Kirkpatrick and B. Selman, Science **264**, 1297 (1994).

[3] A. K. Hartmann and M. Weigt, *Phase Transitions in Combinatorial Optimization Problems* (Wiley-VCH, 2001).

[4] *Special Issue on NP-Hardness and Phase Transitions*, edited by O. Dubois, R. Monasson, B. Selman and R. Zecchina [Theor. Comput. Sci. **265**, 1 (2001)].

[5] R. Monasson and R. Zecchina, Phys. Rev. Lett. **76**, 3881 (2006).

[6] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman and L. Troyansky, Nature (London) **400**, 133 (1999).

[7] M. Ercsey-Ravasz and Z. Toroczkai, Nat. Phys. **7**, 966 (2011).

[8] D. Mitchell, B. Selman and H. Levesque, in *Proceedings of the Tenth National Conference on Artificial Intelligence AAAI-92* (AAAI Press, San Jose, CA, July 1992), p. 456.

[9] S. Cocco and R. Monasson, Phys. Rev. Lett. **86**, 1654 (2001).

[10] C. H. Papadimitriou, in *Proceedings of the 32nd Annual Symposium on Foundations of Computer Science* (IEEE Computer Society Press, New York, 1991), p. 163.

[11] S. Seitz, M. Alava and P. Orponen, J. Stat. Mech.: Theory Exp. (2005), P06006.

[12] J. Ardelius and E. Aurell, Phys. Rev. E **74**, 037702 (2006).

[13] M. Alava, J. Ardeliu, E. Aurell, P. Kaski and S. Krishnamurthy *et al.*, Proc. Natl. Acad. Sci. USA **105**, 15253 (2007).

[14] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University Press, USA, 1999).

[15] K. Binder and D. W. Heerman, *Monte Carlo Simulation in Statistical Physics*, 2nd ed. (Springer-Verlag, Berlin, 1992).

[16] S. H. Lee, M. Ha, C. Jeon and H. Jeong, Phys. Rev. E **82**, 061109 (2010).

[17] M. Mézard, G. Parisi and R. Zecchina, Science **297**, 812 (2002).

[18] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian and L. Zdeborová, Proc. Natl. Acad. Sci. USA **104**, 10318 (2007).

[19] W. Barthel, A. K. Hartmann, M. Leone, F. Ricci-Tersenghi and M. Weigt *et al.*, Phys. Rev. Lett. **88**, 188701 (2002).

[20] K. Li, H. Ma and H. Zhou, Phys. Rev. E **79**, 031102 (2009).